JOURNAL OF COMPUTER AND SYSTEM SCIENCES 7, 579–582 (1973)

A Characterization Theorem for n-Parallel Right Linear Languages*

ROBERT D. ROSEBRUGH

Mathematics Department, Dalhousie University, Halifax, Nova Scotia, Canada

AND

DERICK WOOD

Department of Applied Mathematics, McMaster University, Hamilton, Ontario, Canada

A characterization of the family of n-parallel right linear languages by nondeterministic generalized sequential machines with accepting states is proved.

INTRODUCTION

The family of n-parallel right linear languages (n-PRLL), studied in Rosebrugh (1972) and Wood (1972), is a proper subfamily of the n-right linear simple matrix languages (n-RLSML) studied in Ibarra (1970). However, the family of n-PRLL are of independent interest because they capture the essence of the parallel processing of a sentential form in a grammar. This is demonstrated by the result proved in this paper, that the family of n-PRLL is the smallest family of languages containing the prototype language $\{a_1{}^i \cdots a_n{}^i : i \geqslant 1\}$ and closed under a-NGSM maps.

BASIC NOTATION

An *alphabet* is a nonempty finite set.

For $n \geqslant 1$, *an n-parallel right linear grammar* (n-PRLG) is an $(n + 3)$-tuple $G = (N_1, ..., N_n, T, S, P)$, where N_i, $1 \leqslant i \leqslant n$, are mutually disjoint nonterminal alphabets, T is a *terminal* alphabet, S not in N is a *sentence symbol*, where $N = N_1 \cup \cdots \cup N_n$, and $P \subseteq N \times (T^*N \cup T^*) \cup \{S\} \times (N_1 \cdots N_n \cup T^*)$ is a finite set of rules, written $X \rightarrow x$. A rule $S \rightarrow x$, x in T^* is *trivial*, all other rules are *nontrivial*.

The yield relation is defined as follows: for x, y in $(N \cup T \cup \{S\})^*$, $x \underset{G}{\Rightarrow} y$ iff

* Work carried out by the first author under a National Research Council of Canada Bursary and by the second author under a National Research Council of Canada Grant No. A-7700.

either $x = S$ and $S \rightarrow y$ is in P or $x = y_1X_1 \cdots y_nX_n$, $y = y_1x_1 \cdots y_nx_n$, where y_i in T^*, x_i in $T^*N \cup T^*$, X_i in N_i, and $X_i \rightarrow x_i$ in P, $1 \leqslant i \leqslant n$. If x, y in $(N \cup T \cup \{S\})^*$ and $l > 0$, then $x \underset{G}{\Rightarrow}^l y$ iff there exists a sequence D: $x_0 \underset{G}{\Rightarrow} x_1 \underset{G}{\Rightarrow} \cdots \underset{G}{\Rightarrow} x_l$, $x_0 = x$, $x_l = y$ and x_i in $(N \cup T \cup \{S\})^*$, $0 \leqslant i \leqslant l$. Then D is an *l-step G-derivation sequence*. Similarly $x \underset{G}{\Rightarrow}^+ y$, a *$G$-derivation*, iff there exists $l > 0$ such that $x \underset{G}{\Rightarrow}^l y$ and $x \underset{G}{\Rightarrow}^* y$ iff $x \underset{G}{\Rightarrow}^+ y$ or $x = y$.

The language generated by an n-PRLG is $\{x: S \underset{G}{\Rightarrow}^+ x,\ x \text{ in } T^*\}$ denoted $L(G)$. A language $L \subseteq T^*$ is an *n-parallel right linear language* (N-PRLL) iff there exists an n-PRLG G such that $L = L(G)$. The family of n-PRLL is denoted by $\mathscr{R}_n^\epsilon$.

A *nondeterministic generalized sequential machine with accepting states* (a-NGSM) is a 6-tuple $M = (Q, T, \Delta, \delta, q_0, F)$ where Q is a *state* alphabet, T is an *input* alphabet, Δ is an *ouput* alphabet, $\delta: Q \times T \rightarrow 2^{Q \times \Delta^*}$, finite subsets only, is the *state transition function*, q_0 in Q is the *initial* state and $F \subseteq Q$ is a set of *final* states.

In the usual way the map δ can be extended to give the map $\delta^*: Q \times T^* \rightarrow 2^{Q \times \Delta^*}$. If x is in T^* then $M(x) = \{z: (q, z) \text{ in } \delta^*(q_0, x), \text{ for some } q \text{ in } F, z \text{ in } \Delta^*\}$ and if $L \subseteq T^*$, then $M(L) = \bigcup_{x \text{ in } L} M(x)$.

Characterization of $\mathscr{R}_n^\epsilon$

Definition. For $n \geqslant 1$, let $T_n = \{a_1, \ldots, a_n\}$ be an alphabet of n distinct symbols. For $n \geqslant 1$, let $L_n = \{a_1^i a_2^i \cdots a_n^i : i \geqslant 1\} \subseteq T_n^*$, it is clear that L_n is in $\mathscr{R}_n^\epsilon$.

Theorem. *Characterization theorem for $\mathscr{R}_n^\epsilon$. For $n > 0$, $\mathscr{R}_n^\epsilon$ is the smallest family of languages containing L_n and closed under a-NGSM maps.*

Proof. We first show, by construction, that $\mathscr{R}_n^\epsilon$ is closed under a-NGSM maps. Let L be in $\mathscr{R}_n^\epsilon$, $G = (N_1, \ldots, N_n, T, S, P)$ be an n-PRLG for L and $M = (Q, T, \Delta, \delta, q_0, F)$ be an a-NGSM. We give an n-PRLG G', for $M(L)$ which shows that $M(L)$ is in $\mathscr{R}_n^\epsilon$.

Let $G' = (N_1', \ldots, N_n', \Delta, S, P')$, where $N_i' = (Q \times N_i \times Q)$, $1 \leqslant i \leqslant n-1$, $N_n' = Q \times N_n$ and P' contains the following rules:

(1) $S \rightarrow z$ if $S \rightarrow x$ in P and (q, z) in $\delta^*(q_0, x)$ for some q in F,

(2) $S \rightarrow [q_0, X_1, q_1][q_1, X_2, q_2] \cdots [q_{n-1}, X_n]$ for all sequences $q_0, \ldots, q_{n-1}$ of members of Q, if $S \rightarrow X_1 \cdots X_n$ is in P, X_i in N_i, $1 \leqslant i \leqslant n$,

(3) $[q_i, X_j, q_k] \rightarrow z[q, Y_j, q_k]$ if $X_j \rightarrow yY_j$ is in P, y in T^*, X_j, Y_j in N_j, q_i, q_k in Q, $1 \leqslant j < n$, and (q, z) in $\delta^*(q_i, y)$,

(4) $[q_i, X_n] \rightarrow z[q, Y_n]$ if $X_n \rightarrow yY_n$ is in P, where (q, z) in $\delta^*(q_i, y)$,

(5) $[q_i, X_j, q_k] \rightarrow z$ if $X_j \rightarrow x$ is in P, x in T^*, X_j in N_j and (q_k, z) in $\delta^*(q_i, x)$, $1 \leqslant j < n$, and

(6) $[q_i, X_n] \rightarrow z$ if $X_n \rightarrow x$ is in P, x in T^* and (q, z) in $\delta^*(q_i, x)$, where q is in F.

G' generates all $M(x)$ for all x in T^* generated trivially by G (rules of type (1), above). If a word x is generated nontrivially by G, each word in $M(x)$ is generated by G', which deposits the "translation" of a word deposited by G. A G'-derivation sequence keeps track of the state transitions of M in the first component of the nonterminals from N'. The third component of the nonterminals in N' is used to match states at the boundaries corresponding to a factorization of the word according to the nonterminal from which it is generated.

Therefore $M(L) = L(G')$.

Secondly, again by construction, we show that every n-PRLL L is the image of L_n under some a-NGSM map.

Let $G = (N_1, \ldots, N_n, T, S, P)$ be an n-PRLG for L. We will now construct an a-NGSM $M = (Q, T_n, T, \delta, q_0, F)$ such that $L = M(L_n)$. Number the initial rules of G letting the first k be the nontrivial rules and the trivial rules be numbered from $k + 1$ to m. Let $Q = N \times \{1, \ldots, k\} \cup \{q_0, q_{k+1}, \ldots, q_m, q_f\}$, $F = \{q_{k+1}, \ldots, q_m, q_f\}$ and specify the map δ as follows:

(1) $$\begin{aligned}\delta(q_0, a_1) = {} & \{(q_i, x)\colon x \text{ in } T^*,\ S \rightarrow x \text{ in } P \text{ is the } i\text{th rule}\} \\ & \cup \{([Y_1, j], y_1)\colon S \rightarrow X_1 \cdots X_n \text{ in } P \text{ is the } j\text{th rule} \\ & \quad \text{and } X_1 \rightarrow y_1 Y_1 \text{ is in } P,\ y_1 \text{ in } T^*\} \\ & \cup \{([X_2, j], y_1)\colon S \rightarrow X_1 \cdots X_n \text{ in } P \text{ is the } j\text{th rule} \\ & \quad \text{and } X_1 \rightarrow y_1 \text{ is in } P,\ y_1 \text{ in } T^*\},\end{aligned}$$

(2) $$\delta(q_i, a_j) = \{(q_i, \epsilon)\},\ k + 1 \leqslant i \leqslant m,\ 1 \leqslant j \leqslant n,$$

(3) $$\begin{aligned}\delta([Y_i, j], a_i) = {} & \{([Z_i, j], z_i)\colon Y_i \rightarrow z_i Z_i \text{ is in } P,\ Y_i, Z_i \text{ in } N_i,\ z_i \text{ in } T^*\} \\ & \cup \{([X_{i+1}, j], z_i)\colon Y_i \rightarrow z_i \text{ is in } P,\ Y_i \text{ in } N_i,\ z_i \text{ in } T^*, \\ & \quad X_{i+1} \text{ is the } (i+1)\text{th nonterminal in the } j\text{th initial rule}\}, \\ & \quad \text{for } 1 \leqslant j \leqslant k,\ 1 \leqslant i \leqslant n - 1,\end{aligned}$$

(4) $$\begin{aligned}\delta([Y_n, j], a_n) = {} & \{([Z_n, j], z_n)\colon Y_n \rightarrow z_n Z_n \text{ is in } P,\ z_n \text{ in } T^*\} \\ & \cup \{(q_f, z_n)\colon Y_n \rightarrow z_n \text{ is in } P,\ z_n \text{ in } T^*\}, \text{ and}\end{aligned}$$

(5) $\delta(q, a) = \varnothing$ otherwise for all q in Q, a in T_n. The a-NGSM operates in two distinct ways.

(a) M outputs the result of a trivial derivation and reads the remainder of of the input word in a final state with no further outputs (points 1 and 2).

(b) The states of M are used to keep track of a nonterminal from N_i in their first component and which initial rule was used in the second component. M

simulates an n-parallel G-derivation by processing the n-positions sequentially, that is, all the derivations in the first position, followed by all the derivations in the second position, and so on. The number of steps to be carried out in each position is controlled by an input word $a_1^p \cdots a_n^p$, $p \geqslant 1$. Whenever an a_i is input a derivation step takes place in the ith position. Note that M can only proceed if the input symbol a_i and the present state $[X_l, k]$ where X_l in N_l, are such that $i = l$.

It can be seen that x is in $M(L_n)$ iff x is in $L(G)$. Therefore $L = M(L_n)$.

Finally, let $\mathscr{F}_n$ be the smallest family of languages containing L_n and closed under a-NGSM maps. Then $\mathscr{F}_n \subseteq \mathscr{R}_n^\epsilon$ by the first part of this proof. Further by the second part of this proof, since for all L in $\mathscr{R}_n^\epsilon$, $L = M(L_n)$ for some a-NGSM M, we have $\mathscr{R}_n^\epsilon \subseteq \mathscr{F}_n$. Hence the result.

Acknowledgment

The authors thank the referee for his detailed comments which have improved the format of this paper.

References

1. O. Ibarra, Simple matrix languages, IC 17,4 (1970), pp. 359–394.
2. R. D. Rosebrugh, Restricted parallelism and regular grammars, M.Sc. Dissertation, McMaster University, 1972.
3. D. Wood, n-linear simple matrix languages and n-parallel linear languages (unpublished).